ESTIMATION SÉQUENTIELLE (H2)

La **théorie de l'estimation séquentielle** est un exemple de **problème de décision séquentielle** (cf aussi **théorie de l'estimation**, **théorie séquentielle**).

En effet, dans un problème d'**analyse séquentielle** (estimation, test, etc), la taille N de l'**échantillon** observé n'est pas donnée a priori mais résulte :

(a) de l'**observation** du **phénomène** considéré. En général, l'observation s'effectue au cours du temps : **expérimentation** progressive, **sondage** ou **échantillonnage**, etc ;

(b) de conditions fixées initialement : exigence de **précision**, **seuil de réponse** (réaction à un **stimulus**, etc), nombre ou **proportion empirique** de pannes observées ou de défectueux constatés, etc.

Autrement dit, le **statisticien** continue l'observation (**expérience**, **sondage** ou **échantillonnage**, etc) jusqu'à ce que des conditions données au préalable soient réalisées, ce qui définit une **règle d'arrêt**. Dans ce cadre, N devient donc aléatoire et la **procédure** doit être adaptée à cette **situation statistique**.

(i) Dans le contexte de la théorie de la décision séquentielle, on dit que :

(a) une décision séquentielle (pure) $\delta = (\delta_n)_{n \in \mathbf{N}^*}$ constitue un **estimateur séquentiel** du **paramètre** (non transformé) $\theta \in \Theta$ ssi tous les **espaces de décision** D_n ($n \in \mathbf{N}^*$) sont inclus dans Θ ;

(b) δ est un estimateur séquentiel du paramètre (transformé) $\tau = g(\theta)$ ssi tous les espaces D_n sont inclus dans $\mathbf{R}^Q$ (avec $g : \Theta \mapsto \mathbf{R}^Q$).

Le problème de décision séquentielle ainsi défini est appelé **problème d'estimation séquentielle**.

(ii) Deux principales procédures sont mises en œuvre :

(a) soit, lorsque c'est possible, on augmente la taille N de l'échantillon jusqu'à ce que l'estimateur considéré possède une **précision** donnée, appréciée eg :

$(a)_1$ par un **écart quadratique moyen** (ou une autre **fonction de perte**) lorsque sa valeur est minimum (cas d'un **estimateur ponctuel**) ;

$(a)_2$ par une **région de confiance minimale** pour un **seuil de confiance** donné (cas d'un **estimateur ensembliste**) (cf **coefficient de confiance**) ;

(b) soit on minimise la **fonction de risque** associée à un estimateur pour un coût donné (cf **fonction de coût**) : budget alloué à une expérience, à un sondage ou à un échantillonnage, etc.

(iii) Si N^* désigne la taille (aléatoire) de l'échantillon prélevé jusqu'à arrêt de l'observation du **phénomène** étudié, la théorie conduit à examiner deux cas :

(a) étude de la **loi** de N^* à distance finie (ie lorsque $N^* < +\infty$). Cette loi est souvent difficile à déterminer ;

(b) étude de la **loi asymptotique** de N^* (ie étude de la convergence de sa loi lorsque $N^* \to +\infty$). Dans le cas où $\Theta = \mathbf{R}$, et sous des conditions générales, le **théorème de ANSCOMBE** justifie l'utilisation des **écarts-types** calculés à partir d'un échantillon X de grande taille (quoique à distance finie) $N >> 0$, comme si N n'était pas aléatoire.